MESSAGE ERROR ANALYSIS OF LOOPY BELIEF PROPAGATION

Xiangqiong Shi, Dan Schonfeld, Daniela Tuninetti

Electrical and Computer Engineering
University of Illinois at Chicago
{ xshi4,dans,danielat }@uic.edu

ABSTRACT
The loopy belief propagation algorithm (LBP) is known to perform extremely well in many practical problems of probability inference and learning on graphical models, even in presence of multiple loops. Although general necessary conditions for convergence of LBP to a unique fixed point solution are still unknown, various techniques have been explored to understand error propagation when LBP fails to converge. In this paper, we rely on the contractive mapping of message errors to present novel distance bounds between multiple fixed point solutions when LBP does not converge. We give examples of networks where our bounds are tighter than existing ones.

Index Terms— Graphical Model, Bayesian Networks, Markov Random Fields, Loopy Belief Propagation, Error Analysis

1. INTRODUCTION
Probabilistic inference for large-scale multivariate random variables is very computationally expensive. Belief propagation algorithms (BP) are designed to reduce the computational burden by exploiting the factorization of joint density functions captured by the topological structure of graphical models [1]. BP is known to converge to the correct solution of graphs that are a tree or contain a single loop [2]. BP for graphs with loops is an iterative method referred to as loopy belief propagation (LBP). General necessary convergence conditions of LBP to a unique fixed-point solution are still unknown, though sufficient conditions are investigated extensively in the past [3, 4, 5, 6]. Message error propagation when LBP fails to converge has been studied to understand how the quality of convergence will be affected [3].

Tatikonda and Jordan [5] related the convergence of LBP to the uniqueness of a sequence of Gibbs measures defined on the associated computation tree and developed testable sufficient convergence conditions. Heskes [6] developed sufficient conditions for uniqueness of fixed points of LBP by ensuring unique minima of the Bethe free energy. Ihler et al.[3] analyzed the contractive effect of message error propagation in belief networks using dynamic range measure as a metric, and obtained a sufficient convergence condition and distance bounds between LBP beliefs. Based on contractive mapping of message propagation as well, Mooij and Kappen [4] derived sufficient conditions for convergence of LBP, which are shown to be valid for potential functions containing zeros.

In this paper, we focus on the actual errors on the beliefs when LBP fails to converge. We extend the mathematical framework of [3] by introducing a \textit{maximum error measure} as an error metric, and derive much novel upper- and lower-bounds on error propagation. Our bounds are tighter than existing ones in many instances. In this paper we limit ourselves to error propagation bounds and do not investigate whether the reached fixed point is indeed the exact solution.

2. ERROR PROPAGATION FOR THE SUM-PRODUCT ALGORITHM
Belief propagation (BP) originated from exact inference on tree structured graphical models, though for graphs with loops it perform approximate inference well. In case of marginalization of global distribution, BP implements a sum-product algorithm, while for Maximum A Posteriori (MAP) it is realized as a max-product algorithm. In the following, we will focus on the sum-product algorithm for graphs with loops.

2.1. LBP Updates
Let us consider a general graphical model $G(V, E)$ whose distribution factors as follows:

$$ p(X) = \frac{1}{Z} \prod_{(s,t) \in E} \psi_{st}(x_s, x_t) \prod_{s \in V} \psi_s(x_s), \quad (1) $$

where $V$ is the set of vertices, $E$ is the set of edges, $Z$ is a normalization factor, $\psi_{st}(x_s, x_t)$ is the pairwise potential function between random variables $x_s$ and $x_t$, and $\psi_s(x_s)$ is the single node potential function on $x_s$. We assume that all the potential functions are positive. The updating rule of the sum-product algorithm for the message sent by node $t$ to its neighbor node $s$ at iteration $i$ is:

$$ m_{ts}^i(x_s) \propto \int \psi_{ts}(x_t, x_s) \psi_t(x_t) \prod_{u \in \Gamma_t \setminus s} m_{us}^{i-1}(x_t) dx_t, \quad (2) $$. 


where $\Gamma_t$ is the set of neighbors of node $t$. The belief, or pseudo-marginal probability of $x_t$, on node $t$ at iteration $i$, is:

$$M_t^i(x_t) \propto \psi_t(x_t) \prod_{u \in \Gamma_t} m_{ut}^i(x_t).$$

(3)

A stable fixed point has been reached if $m_{ts}^i(x_s) = m_{ts}^{i+1}(x_s)$.

For synchronous BP, each iteration of (2) and (3) corresponds to a level in the computation tree [3]. Here we only discuss synchronous BP.

Fig. 1. Graphical models: (a) message passing in a belief network; (b) symmetric graph; and (c) lattice graph.

2.2. Dynamic Range Measure

Various approaches have been presented to derive convergence conditions for the sum-product algorithm, including contraction of message errors on belief networks. Define the message error as a multiplicative functions $e_{ts}^i(x_s)$ that perturbs the fixed point message $m_{ts}^i(x_s)$ into $\hat{m}_{ts}^i(x_s) = m_{ts}^i(x_s)e_{ts}^i(x_s)$ at iteration $i$. Here, we deal with normalized messages. We define incoming message products as $M_{ts}^i(x_t) \propto \psi_t(x_t) \prod_{u \in \Gamma_t \setminus \{s\}} m_{ut}^i(x_t)$ and incoming error products as $E_{ts}^i(x_t) = \prod_{u \in \Gamma_t \setminus \{s\}} e_{ts}^i(x_t)$. Thus, the outgoing message error from node $t$ to node $s$ at iteration $i + 1$ is:

$$e_{ts}^{i+1}(x_s) = \frac{\hat{m}_{ts}^{i+1}(x_s)}{m_{ts}^i(x_s)} = \frac{\int \psi_t(x_t, x_s) M_{ts}^i(x_t) E_{ts}^i(x_t) dx_t}{\int \psi_t(x_t, x_s) M_{ts}^i(x_t) E_{ts}^i(x_t) dx_t dx_s} \times \frac{\int \psi_t(x_t, x_s) M_{ts}^i(x_t) dx_t}{\int \psi_t(x_t, x_s) M_{ts}^i(x_t) dx_t}.$$

The dynamic range measure of the error introduced by Ihler et al. [3] is defined as:

$$d(e_{ts}) = \max_{a,b} \sqrt{e_{ts}(a)/e_{ts}(b)},$$

(4)

where $d(e_{ts}) \rightarrow 1$ when $e_{ts} \rightarrow 1$. In [3, Th.8] it was shown that when $d(\psi_t) = \max_{a,b,c,d} \sqrt{\psi_t(a,b)/\psi_t(c,d)}$ is finite, the dynamic range satisfies the following contraction:

$$d(e_{ts}^{i+1}) \leq \frac{d(\psi_t)^2 d(E_{ts}^i)}{d(\psi_t)^2 + d(E_{ts}^i)},$$

(5)

i.e., the dynamic range of the outgoing message error is bounded by the dynamic range of the potential function and the dynamic range of the incoming error product.

2.3. Maximum Error Measure

To study the characteristics of message error propagation, dealing directly with errors is more interesting than dealing with dynamic range. We thus introduce the following maximum multiplicative error function as an error measure:

$$\max_{x_s} e_{ts}^{i+1}(x_s) = \max_{x_s} \frac{\int \psi_t(x_t, x_s) M_{ts}^i(x_t) E_{ts}^i(x_t) dx_t}{\int \psi_t(x_t, x_s) M_{ts}^i(x_t) dx_t} \times \frac{\int \psi_t(x_t) M_{ts}^i(x_t) dx_t}{\int \psi_t(x_t, x_s) M_{ts}^i(x_t) dx_t},$$

where $\psi_t(x_t) = \int \psi_t(x_t, x_s) dx_s$. It is immediate that the maximum error measure approaches one when errors vanish.

In the following, we use the maximum multiplicative error to prove that message errors are bounded and to derive an upper-bound on the distances between fixed points of beliefs.

2.4. Error Bounds

We shall omit reference to the iteration number of the messages and errors for simplicity and clarity of the presentation.

Theorem 1. Multiplicative outgoing errors are bounded as:

$$\left( \frac{d(\psi_t) d(\psi_t) d(E_{ts})}{d(\psi_t) d(\psi_t) + d(E_{ts})} \right)^2 \leq \min_{x_s} e_{ts}(x_s) \leq e_{ts}(x_s),$$

$$\leq \max_{x_s} e_{ts}(x_s) \leq \left( \frac{d(\psi_t) d(\psi_t) d(E_{ts})}{d(\psi_t) d(\psi_t) + d(E_{ts})} \right)^2.$$

Proof: By [3, Lemma 26], max $e_{ts}(x_s)$ is maximum when $\psi_t(x_t, x_s) = 1 + (d(\psi_t^2 - 1) \chi_\psi(x_t), \psi_t(x_t) = 1 + (d(\psi_t^2 - 1) \chi_\psi(x_t))$ and $E_{ts} = 1 + (d(E_{ts}^2) - 1) \chi_E(x_t)$, where $\chi_\psi$, $\chi_\psi$, and $\chi_E$ are indicator functions. Define the quantities:

$$M_A = \int M_{ts}(x_t) \chi_\psi(x_t) dx_t,$n
$$M_B = \int M_{ts}(x_t) \chi_E(x_t) dx_t,$n
$$M_C = \int M_{ts}(x_t) \chi_E(x_t) dx_t,$n
$$M_{AE} = \int M_{ts}(x_t) \chi_\psi(x_t) \chi_E(x_t) dx_t,$n
$$M_{BE} = \int M_{ts}(x_t) \chi_\psi(x_t) \chi_E(x_t) dx_t,$n

and $\alpha_1 = d(\psi_t)^2 - 1$, $\alpha_2 = d(\psi_t)^2 - 1$, and $\beta = d(E_{ts})^2 - 1$. The maximum multiplicative error is upper-bounded by max $e_{ts}(x_s) \leq \Delta_1$ where

$$\Delta_1 = \max_{\{M\}} \frac{1 + \alpha_1 M_A + \beta M_B + \alpha_1 M_{AE}}{1 + \alpha_2 M_B}.$$

The maximum over $\{M\}$ is obtained when $M_{AE} = M_A = M_E = 1 - M_B$ and $M_{BE} = 0$, which gives

$$\Delta_1 = \max_{M_E} \frac{1 + (\alpha_1 + \beta + \alpha_1 \beta) M_E}{1 + \alpha_2 - \alpha_2 M_E}.$$

Taking the derivative wrt $M_E$ and setting it to zero, we obtain

$$\max_{x_s} e_{ts}(x_s) \leq \Delta_1 = \left( \frac{d(\psi_t) d(\psi_t) d(E_{ts})}{d(\psi_t) d(\psi_t) + d(E_{ts})} \right)^2.$$

(6)
Similarly to what we have done so far, we can minimize \( \min_{x_s} e_{ts}(x_s) \) with respect to \( \psi_{ts}(x_t, x_s), \psi_{ts}(x_t) \) and \( E_{ts}(x_t) \), to obtain
\[
\min_{x_s} e_{ts}(x_s) \geq \left( \frac{d(\psi_{ts})d(\psi_{ts}) + d(E_{ts})}{d(\psi_{ts})d(\psi_{ts}) + 1} \right)^2 = \frac{1}{\Delta_1}.
\]  
\[
(7)
\]

**Lemma 1.** The upper bound on the multiplicative error provided in Theorem 1 is tighter than the following upper bound from [3, Th.2 and Th.8]:
\[
\max_{x_s} e_{ts}(x_s) \leq d(e_{ts})^2 \leq \left( \frac{d(\psi_{ts})^2d(E_{ts}) + 1}{d(\psi_{ts})^2d(E_{ts})} \right)^2 = \Delta_2.
\]  
\[
(8)
\]

Proof. Since \( \Delta_1 \) in (6) is increasing in \( d(\psi_{ts}) \) we conclude that (6) implies (8), i.e., \( \Delta_1 \leq \Delta_2 \), because
\[
d(\psi_{ts}) = \max_{a,b} \sqrt{\frac{\psi_{ts}(a)}{\psi_{ts}(b)}} = \max_{a,b} \sqrt{\int_{\psi_{ts}(a, x_s)dx_s}\int_{\psi_{ts}(b, x_s)dx_s}} \leq \max_{a,b,c,d} \sqrt{\frac{\psi_{ts}(a, c)}{\psi_{ts}(b, d)}} = \max_{a,b,c,d} \sqrt{\psi_{ts}(a, c)} \psi_{ts}(b, d) = d(\psi_{ts}).
\]

We are interested to know how beliefs will vary at each iteration, when LBP fails to converge on a graphical model. In the following, we will present our uniform distance bound and non-uniform distance bound on fixed points of beliefs.

**Corollary 1.** *(Uniform Distance Bound)*
The log-distance bound of fixed points of belief at node \( s \) is
\[
\sum_{t \in \Gamma_s} \log\left( \frac{d(\psi_{ts})d(\psi_{ts}) + 1}{d(\psi_{ts})d(\psi_{ts}) + 1} \right)^2 = \log G_{sp}(\log \varepsilon) = \log \Delta_3(\varepsilon) - \log \varepsilon, \varepsilon \geq 1.
\]

Proof. Let the function \( \Delta_{ut}(x) = \frac{d(\psi_{ut})d(\psi_{ut}) + 1}{d(\psi_{ut})d(\psi_{ut}) + 1} \), \( x \geq 1 \). Therefore,
\[
d(E_{ts}^i) \leq \prod_{u \in \Gamma_t \setminus s} d(e_{ut}^i) = \prod_{u \in \Gamma_t \setminus s} \max_{x_t} \sqrt{e_{ut}^i(x_t)} = \prod_{u \in \Gamma_t \setminus s} \min_{x_t} \sqrt{e_{ut}^i(x_t)} \leq \varepsilon_{ts}^i = \prod_{u \in \Gamma_t \setminus s} \Delta_{ut}(d(E_{ut}^{i-1})).
\]

Thus, we have
\[
\max_{x_s} E_{sp}^{i+1}(x_s) \leq \prod_{t \in \Gamma_s \setminus p} \max_{x_s} e_{ts}^{i+1}(x_s) \leq \varepsilon_{sp}^{i+1} = \prod_{t \in \Gamma_s \setminus p} d(E_{ts}^i) \leq \prod_{t \in \Gamma_s \setminus p} \Delta_{ts}(\varepsilon_{ts}^i) \leq \prod_{t \in \Gamma_s \setminus p} \Delta_{ts}(\max_{x_t} e_{ut}^i) = \Delta_3(\max_{x_t} e_{ut}^i) = \Delta_3(\max_{t \in \Gamma_s \setminus p} e_{ts}^i).
\]

\( \Delta_3(\max_{t \in \Gamma_s \setminus p} e_{ts}^i) \) is an upper-bound on the incoming error product \( E_{sp}^{i+1}(x_s) \) at iteration \( i+1 \), while \( \max_{t \in \Gamma_s \setminus p} e_{ts}^i \) is the maximum of the upper-bounds on the incoming error products \( E_{ts}(x_t), t \in \Gamma_s \setminus p \) at iteration \( i \).

Denoting \( \varepsilon = \max_{t \in \Gamma_s \setminus p} e_{ts}^i \), let us introduce an error bound variation function:
\[
G_{sp}(\log \varepsilon) = \log \Delta_3(\varepsilon) - \log \varepsilon, \varepsilon \geq 1,
\]
which describes variation of error bounds after each iteration. When \( G_{sp}(\log \varepsilon) = 0 \), the log distance bound log \( \varepsilon \) will reach a fixed point. Since \( G_{sp}^{(2)}(\log \varepsilon) < 0 \) for \( \log \varepsilon > 0 \) and \( G_{sp}^{(1)}(\infty) = -1/2 \), \( G_{sp}^{(1)}(\log \varepsilon) \) will decrease until it is equal to \( -1/2 \). Therefore, besides \( \log \varepsilon = 0 \) (zero crossing point), it only has one crossing point, which is a stable fixed point of function \( G_{sp}(\log \varepsilon) \). In other words, once \( \log \varepsilon \) leaves the zero crossing point, it will stay at this point. The crossing point \( \log \varepsilon^* \) can be obtained by solving \( G_{sp}(\log \varepsilon^*) = 0 \).

Since the distance between fixed points of \( b_u(x_s) \) is
\[
\log e_{ts}(x_s) = \log \left( \prod_{t \in \Gamma_s} e_{ts}(x_s) \right) \leq \log \prod_{t \in \Gamma_s} \Delta_{ts}(\varepsilon^*),
\]
the log-distance bound can be obtained by taking the maximum \( \varepsilon^* \).

**Fig. 2.** Plot of the true error variation function and the error bound variation function.

The error bound variation function using \( \Delta_2 \) in (8) will be \( G_{sp}(\log \varepsilon) = \log \prod_{t \in \Gamma_s \setminus p} d(\psi_{ts})^2d(E_{ts}) + 1/2 \), \( \varepsilon \geq 1 \). Since \( G_{sp}(\log \varepsilon) \leq G_{sp}(\log \varepsilon) \), the fixed point of \( G_{sp}(\log \varepsilon) \) must be smaller than that of \( G_{sp}(\log \varepsilon) \). In other words, our log distance bound will be tighter. Let \( G_{sp}^1 = G_{sp}^0 = G_{sp} \). The two functions are illustrated in Fig.2 for the graph in Fig.1 (b) with all the pairwise potential functions being \((0.7, 0.3; 0.3, 0.7)\) and all the single node potentials being \((1, 1)\). The true error variation function in Fig.2 characterizes the difference between the incoming error product at one level and the incoming error product at the upper level of the
computation tree. We can observe that our error bound variation function just envelops the true error variation function.

![Graph](image)

**Fig. 3.** True distance and two uniform distance bounds with various $\eta$ for the graph in Fig. 1 (b).

The previous corollary works well for uniform graphs. However, it will give loose bounds when every loop contains both strong and weak potentials and each node has different topology. Therefore, we present non-uniform distance bound.

**Corollary 2. (Non-uniform Distance Bound)**

The log-distance bound of fixed points of belief at node $s$ after $n \geq 1$ iterations is $\sum_{t \in \Gamma_s} \log \left( \frac{d(\psi_{ts})d(\psi_{st})e^{\alpha_t^s+1}}{d(\psi_{ts})d(\psi_{st})+\varepsilon_{ts}} \right)^2$, where $\varepsilon_{ts}$ is updated by $\log \varepsilon_{ts} = \sum_{u \in \Gamma} \log \left( \frac{d(\psi_{ut})d(\psi_{tu})e^{\alpha_t^u+1}}{d(\psi_{ut})d(\psi_{tu})+\varepsilon_{ut}} \right)^2$ with initial condition $\log \varepsilon_{ut} = \sum_{v \in \Gamma} \log \left( \frac{d(\psi_{vu})d(\psi_{uv})e^{\alpha_t^v+1}}{d(\psi_{vu})d(\psi_{uv})+\varepsilon_{vt}} \right)^2$.

**Proof.** It can be easily proved from Corollary 1.

**2.5. Simulation Results**

Our uniform and non-uniform bounds are derived based on maximum error measure, while those in [3, Th.13, Th.14] are derived based on dynamic range measure. When LBP almost converges, their bounds are tighter than ours; however, when LBP does not converge, our bounds become tighter. In other words, their bounds imply better convergence condition, while our bounds give tighter distance bounds between non-unique fixed points of beliefs.

Assume all the pairwise potential functions are $(\eta, 1-\eta; 1-\eta, \eta)$ where $\eta > 0.5$ and all the single node potentials are $(1, 1)$ for the graphs in Fig. 1(b) and Fig. 1(c). Let us compare our distance bound with [3, Th.13] for the graph in Fig. 1 (b) with various $\eta$. Since the graph is completely uniform, beliefs are the same on each node. In other words, each node has the same distance of fixed points. From the result in Fig. 3, we can see that our distance bounds of fixed points are much tighter when $\eta$ increases. When the maximal log distance of beliefs is zero, LBP converges. Thus, we can obtain the critical value $\eta < 0.66$ for convergence for Fig. 1 (b), which is very close to the empirical value $\eta < 0.67$.

Let us see the performances of non-uniform bounds for the graph in Fig. 1 (c). Fig. 4 shows the distance bounds of fixed points for each node in the graph with various $\eta$. We can see that our error bounds are much tighter when $\eta$ increases. The empirical critical value of $\eta$ for convergence is $\eta < 0.79$. Ihler et al. [3] obtained $\eta < 0.79$, while we get $\eta < 0.76$ to ensure convergence, which is still comparable.

**3. CONCLUSION**

In this paper, we derive tight upper- and lower-bounds on error propagation in synchronous belief networks. We subsequently rely on this bound to provide a uniform distance bound and a non-uniform distance bound for fixed points of beliefs. Future efforts will focus on extension of our analysis to the max-product algorithm.

**4. REFERENCES**


